Clustering the administrative districts for the city of São Paulo (SP, Brazil)

A gastronomic adventure in the times of social isolation

Francisco Martellini

May 5, 2020

Introduction

- Our objective is make a exploratory research in the districts (neighbourhoods) of the São Paulo City
- Maybe we can find similarities that can be useful to someone that have the necessity to stay in home understanding the city.
- The method uses is K-Means Clustering

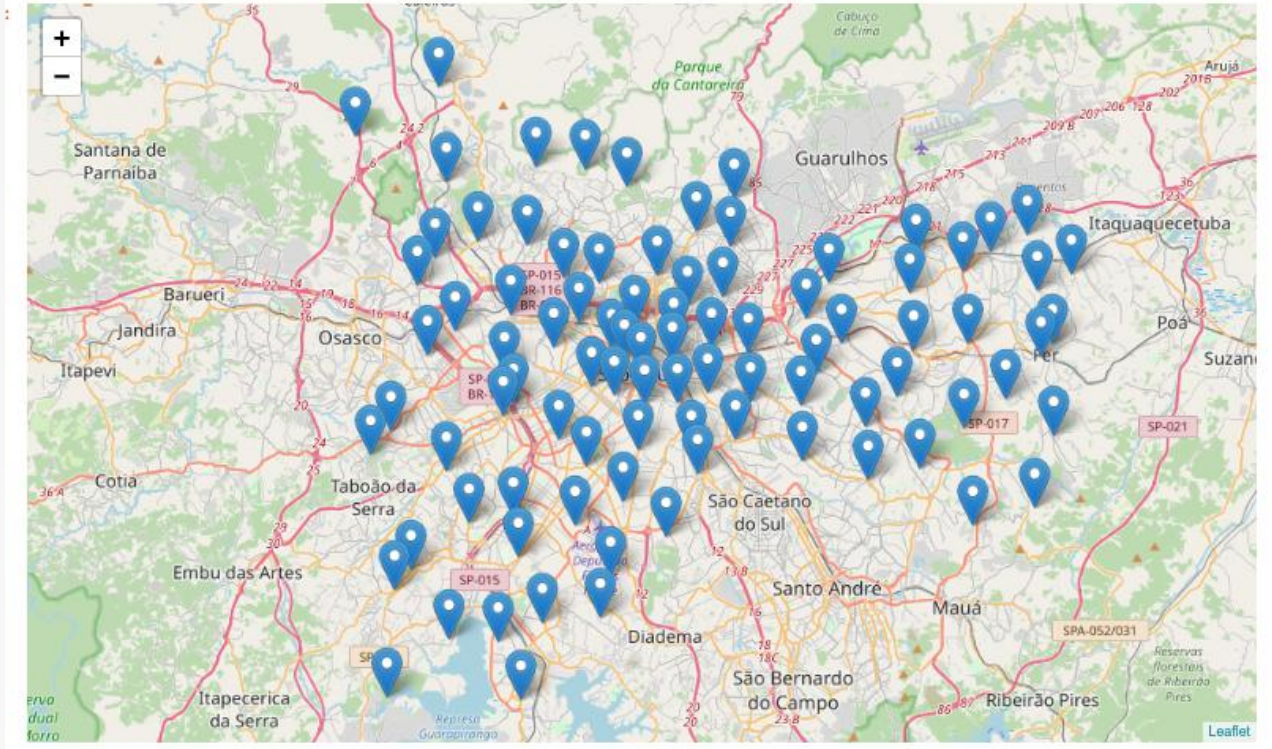
Data Acquistion

- The data was obtained from two sources:
  - The Foursquare places API
  - A Kaggle dataset made by Caio B. Silva, avaiable here:
    - https://www.kaggle.com/caiobsilva/sp-district-coordinates/version/1
    - This dataset have the geolocations for all districts of the São Paulo city and is in conformity with the City Law 11.220/1992

Exploratory data analysis

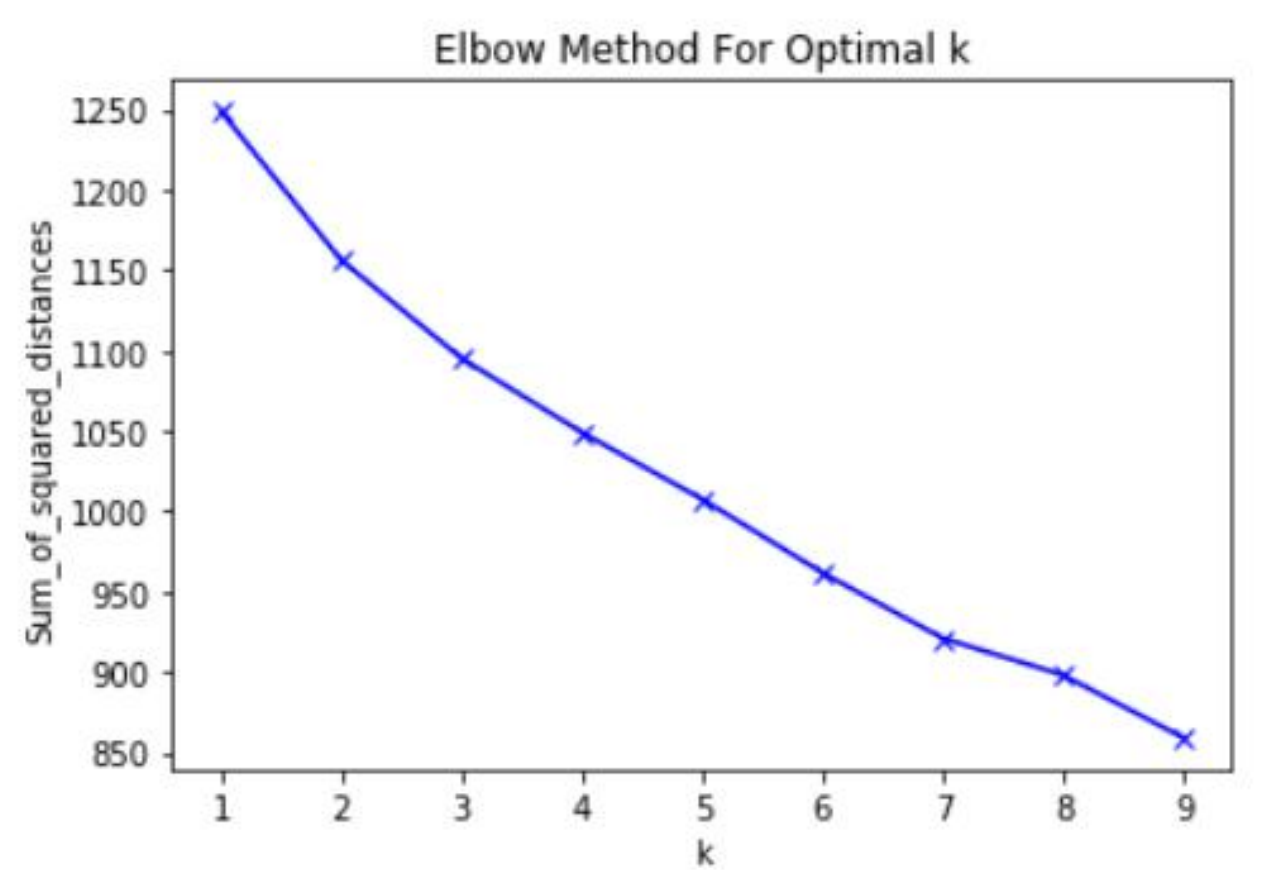
Picture 1: Map of districts for the city of São Paulo and the python code.

```
# create map of Sao Paulo using latitude and longitude values
map_saopaulo = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, district_name, population in zip(spdist_df['Latitude'], spdist_df['Longitude'], spdist_df['Distric
    label = 'Name: {}, Population: {}'.format(district_name,population)
    label = folium.Popup(label, parse_html=True)
    folium.Marker(
        [lat, lng],
        popup=label).add_to(map_saopaulo)

map_saopaulo
```



Predictive modelling

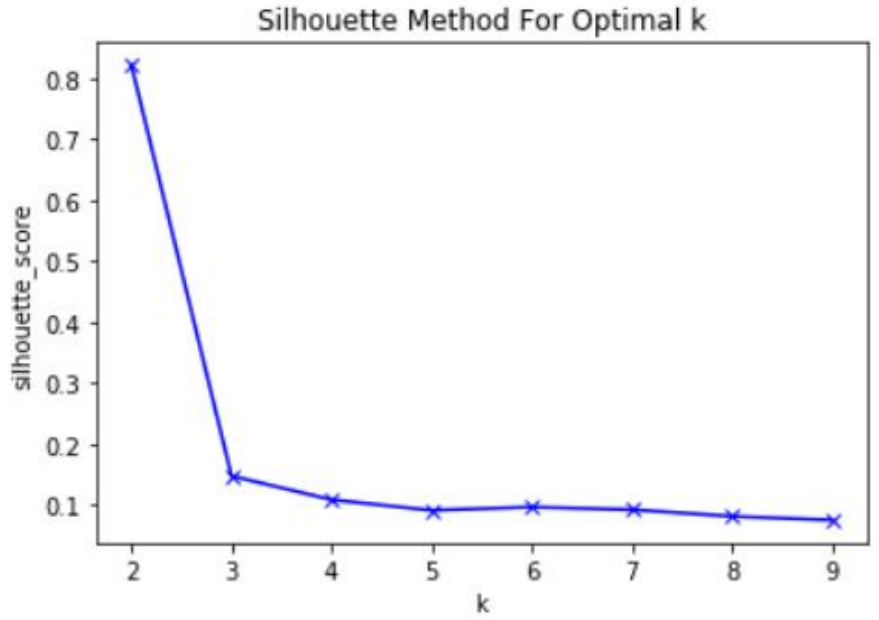
Picture 2: Plot for the Elbow Method to K-Means Cluster.



Predictive modelling

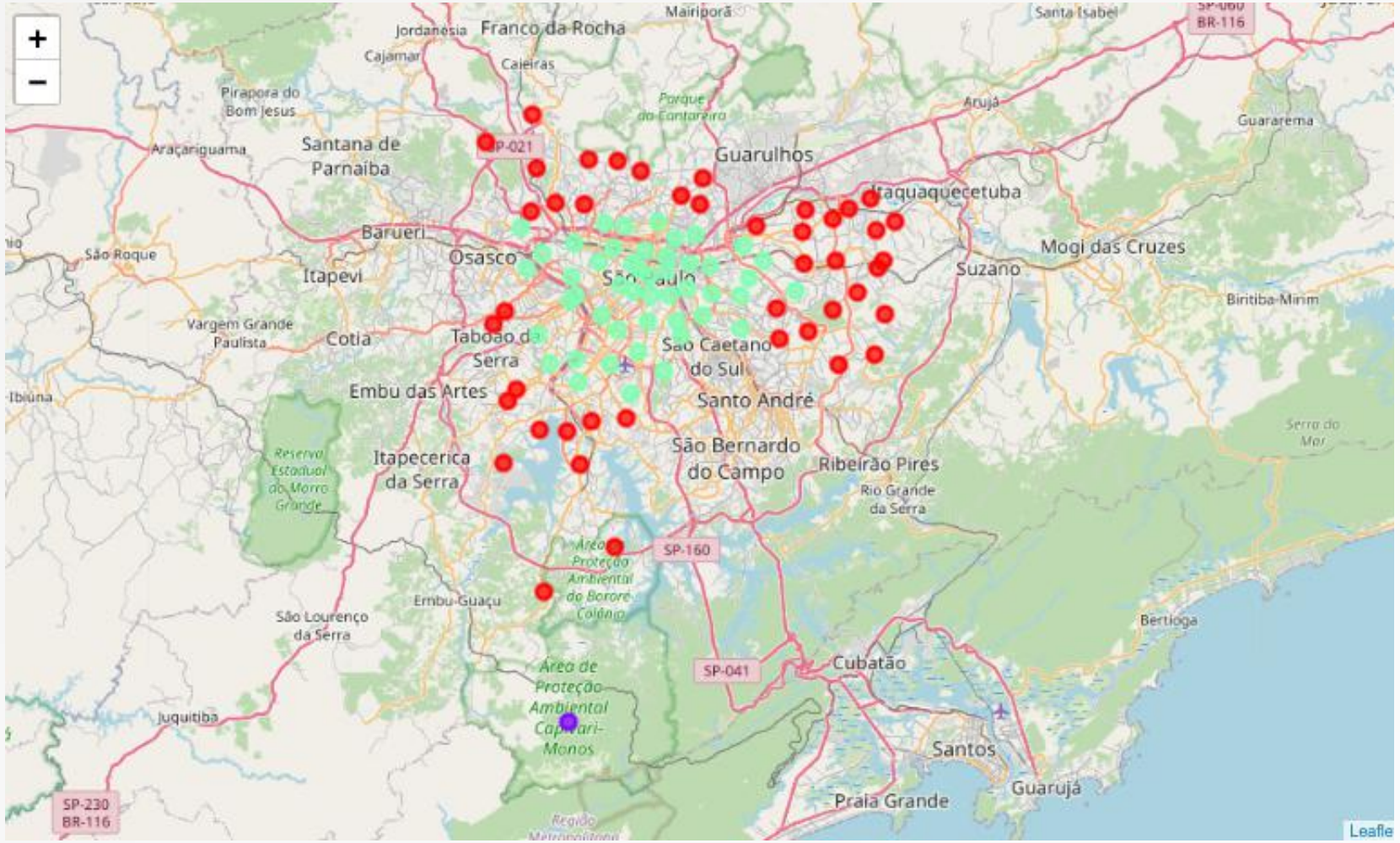
Picture 3: Plot for the Silhouette Method to K-Means Cluster.

```
For n_clusters=2, The Silhouette Coefficient is 0.8210573962604745
For n_clusters=3, The Silhouette Coefficient is 0.1475455275919334
For n_clusters=4, The Silhouette Coefficient is 0.10848908210594288
For n_clusters=5, The Silhouette Coefficient is 0.09116363507628773
For n_clusters=6, The Silhouette Coefficient is 0.09659645843672576
For n_clusters=7, The Silhouette Coefficient is 0.09220412097070717
For n_clusters=8, The Silhouette Coefficient is 0.0813724485647067
For n_clusters=9, The Silhouette Coefficient is 0.07520483576327115
```



Conclusions

Picture 4: Clusters for the districts of São Paulo city.



Conclusions

Cluster 1		**Cluster 2**		**Cluster 3**	
Bakery	44	Brazilian Restaurant	1	Pizza Place	38
Pizza Place	41	Flower Shop	1	Bakery	32
Gym / Fitness Center	41	Flea Market	1	Pet Store	32
Brazilian Restaurant	40	Film Studio	1	Dessert Shop	31
Dessert Shop	30	Field	1	Ice Cream Shop	27
Gym	29	Food & Drink Shop	1	Italian Restaurant	27
Bar	25	Food	1	Burger Joint	24
Japanese Restaurant	20	Zoo	1	Brazilian Restaurant	22
Restaurant	17	Fish & Chips Shop	1	Gym / Fitness Center	22
Pet Store	14	Food Court	1	Bar	18